USING EVIDENCE FEED-FORWARD HIDDEN MARKOV MODELS

Michael S. Del Rose*, Philip Frederick
U.S. Army RDECOM-TARDEC
Warren, MI, 48397-5000

Christian Wagner
Oakland University
Rochester, MI 49306

ABSTRACT

Visual Understanding is an increasing field of research thanks to the advances in image processing, object detection, classification, and advanced computational intelligence techniques. Hidden Markov Models (HMM) are one of these techniques which have been used extensively for this problem. This paper will introduce a new type of HMM, called Evidence Feed Forward Hidden Markov Models, that not only increase the classification rate for sparse messy data, but outlines a whole new theory towards changing the way HMM's are conceived. Data is taken from simulated images of people's actions. Over processing is performed to decrease the likelihood of correct classification. Finally, the over-processed, sparse data is used to train and test the Evidence Feed-Forward HMM and the standard HMM. Results are compared.

1. INTRODUCTION

Visual Understanding (VU) is increasing with the growing advances in technology that require VU algorithms to be taken out of the research labs and into fully developed programs and systems. A sub research area of VU is Visual Human Intent Analysis (VHIA). This area may also be referred to as *visual human behavior identification*, *action or activity recognition*, and *understanding human actions from visual cues*. In static self security systems *visual human behavior identification* systems will aide or replace security guards monitoring CCTV feeds. Television stations will require *activity recognition* systems to automatically categorize and store or quickly search for certain scenes in a database. The U.S. Army is pushing robotics to replace the soldier, thus requiring the need to *understand human actions from visual cues* to determine hostile actions from people so the robot can take appropriate actions to secure itself. These are just a few areas where VHIA will increase current state of the art in the development and use of future systems.

This paper concentrates on new research in the area of Hidden Markov Models (HMM) to the extent of redefining the way HMMs are built. Section 2 will give a background of recent work in the area of Visual Intent Analysis classification. Section 3 will discuss the Evidence Feed Forward Hidden Markov Model as well as provide an example to help illustrate. Section 4 will give the equations for solving the three common HMM problems. Section 5 shows the results of the Evidence Feed-Forward HMM on a problem with over-processed data, and section 6 summarize.

2. BACKROUND

In the area of Visual Intent Analysis classification, there are several research areas. M. Cristani et al [1] uses non-traditional AI methods by taking in both audio and visual data to determine simple events in an office. First they remove foreground objects and segment the images in the sequence. This output is coupled with the audio data and a threshold detection process is used to identify unusual events. These event sequences are put into an audio visual concurrence matrix (AVC) to compare with known AVC events.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
11 MAY 2010	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Using Evidence Feed-Forward Hidden Markov Models	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Michael S. Del Rose; Phillip Frederick Christain Wagner	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **US Army RDECOM-TARDEC 6501 E 11 Mile Rd Warren, MI 48397-5000, USA Oakland University Rochester, MI 49306**	8. PERFORMING ORGANIZATION REPORT NUMBER **20795**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **US Army RDECOM-TARDEC 6501 E 11 Mile Rd Warren, MI 48397-5000, USA**	10. SPONSOR/MONITOR'S ACRONYM(S) **TACOM/TARDEC/RDECOM**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) **20195**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**6**	

Template matching is performed by M. Dimitrijevic et. al. [2]. They developed a template database of actions based on five male and three female people. Each human action is represented by three frames of their 2D silhouette at different stages of the activity: the frame when the person first touches the ground with one of his/her feet, the frame at the midstride of the step, and the end frame when the person finishes touching the ground with the same foot. The three frame sets were taken from seven camera positions. When determining the event, they use a modified Chamfer's distance calculation to match to the template sequences in the database.

Some traditional AI methods include H. Stern et al. [3] who created a prototype fuzzy system for picture understanding of surveillance cameras. His model is split into three parts, pre-processing module, a static object fuzzy system module, and a dynamic temporal fuzzy system module. The static fuzzy system module classifies pre-processed data as a single person, two people, three people, many people, or no people. The dynamic fuzzy system determines the intent of the person based on the temporal movements.

Another common approach is using Grammars to describe the sequence of movements that make up the action. A. Ogale et. al. [4] uses probabilistic context free grammars (PCFG) in short action sequences of a person from video. Body poses are stored as silhouettes which are used in the construction of the PCFG. Pairs of frames are constructed based on their time slot: the pose from frame 1 and 2 are paired, the pose from frame 2 and 3 are paired, and so on. These pairs construct the PCFG for the given action. When testing the algorithm, the same procedure is followed. Comparing the testing data with the trained data is accomplished through Bayes: $P(s_k|p_i) = P(p_i|s_k)P(s_k)/P(p_i)$, where s_k is the k^{th} silhouette and p_i is the i^{th} pose.

There are a number of traditional and non-traditional Hidden Markov Models (HMM) that are used in trying to understand peoples actions based on visual sequences. A few include Yamato et. al. [5] used HMMs to recognize six tennis strokes with a 25x25 mess feature matrix to describe body positions in each frame. Wilson and Bobick [6] use a Parametric Hidden Markov Model (PHMM) to recognize hand gesture. Oliver et. al. [7] developed a method to detect and classify interactions between people using a Coupled Hidden Markov Model (CHMM) based on simulations. Multi-Observation Hidden Markov Models (MOHMM) are discussed in both [8] and [9] from Xaing and Gong for recognizing break points in video content for separation of activities and detect piggybacking of peoples going through a security door, respectively.

3. EVIDENCE FEED-FORWARD HIDDEN MARKOV MODELS INTRODUCTION

Evidence Feed-Forward HMMs are HMMs that involve positive feed-forward from the current observation nodes into the nodes of the future observation in the Hidden Markov Model (HMM). This is more than an extension to HMMs like Parametric HMMS or Hierarchal HMMs because it relaxes the need for complete independence, disregards the rules of causality as suggested by HMM theory, and it provides a link from evidence to evidence that is not through the hidden nodes, which in the strict sense, Markov models current state only depends on the previous states where in the proposed new model, Evidence Feed-Forward HMM, this is no longer the case. However, an Evidence Feed-Forward HMM can still be classified as an HMM since there is a hidden layer, a network of choices, and evidence that is observed. The learning algorithms and the applications are similar to standard HMMs. The difference comes in the interpretation of how a process should model a real world event.

As an example, take the commonly used Weather Example: A person is locked inside a windowless building and would like to know whether it is raining or not outside. The only evidence he has is whether he sees his boss come inside with or without an umbrella. He constructs an HMM to make his decision. Figure 1 shows the hidden layer is represented by the blue nodes Rain (R) and No Rain (NR). The evidence is represented by the yellow nodes of Umbrella (U) or No Umbrella (NU). This example shows that the evidence (observation)

is only dependent on the hidden layer and not vice versa. Also, the hidden layer is only dependent on the previous day's weather (with some probability). However, we have not taken into account the effect of the evidence affecting the next day's evidence and the weather.

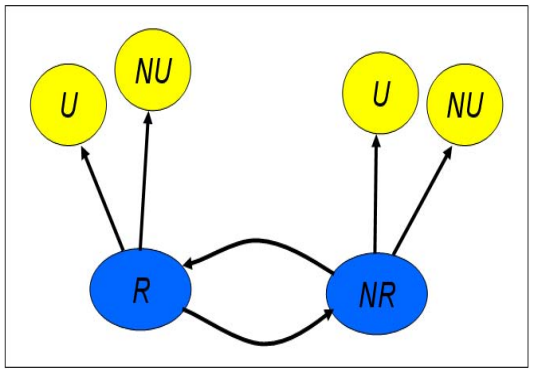

Fig. 1: Weather Example using HMM.

In this example, suppose the boss comes into the building without an umbrella and it is raining. Then, it would be logical to assume that the boss would be more likely to carry an umbrella the next day. This changes the thinking of the evidence portion of HMMs. Previous HMMs assume that the evidence is based only on the current node (hidden) that you are at, so seeing there was an umbrella or not does not have any effect on seeing the next day of an umbrella or not. However, if we look closer at this, we are looking at the actions of the boss as well, so there is a probability associated with his actions (which turn out to be the observations in this HMM). This idea connects the evidence of each event to the evidence of the next event.

By connecting observations to observations, the network gets very complex. However, it can be simplified by assuming that the probability of going to a future observation is only dependent on the probability of the current observation and the current state (hidden) it is in. Applying this to the example, the probability of having an umbrella given it rained the previous day and the boss did not have his umbrella is the same without needing to know the current days weather. I.e. The boss will increase his likelihood by the same amount of carrying his umbrella whether it is raining or not. This does not mean that the likelihood of the boss carrying an umbrella is the same, only the increase is the same. So, if the likelihood of the boss carrying an umbrella is very high compared to not carrying one when it is raining, then this increase will probably not have a large effect on the outcome of the boss not carrying an umbrella when it is raining. See figure 2 for a pictorial view of this example using Evidence Feed-Forward HMMs.

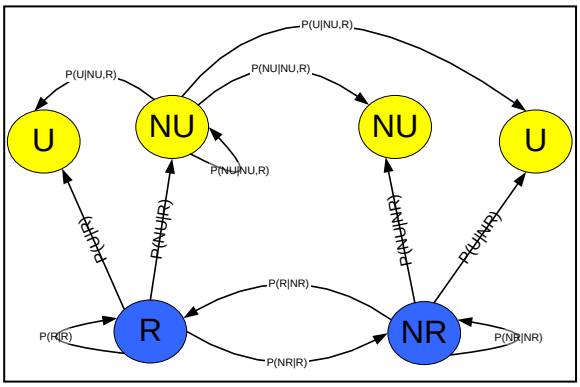

Fig. 2: Weather example using Evidence Feed-Forward Hidden Markov Models.

4. EVIDENCE FEED FORWARD HIDDEN MARKOV MODELS THEORY

Just like standard HMMs, the three common problems an Evidence Feed-Forward HMMs should solve are:

1. Given an observation sequence $O = O_1O_2\ldots O_T$ and a model $\lambda = (A,B,C,\pi)$, compute the probability of the observation sequence given the model ($P(O|\lambda)$).

2. Given the observation sequence O and the model λ, find the optimal path through the hidden state sequence $Q = q_1q_2\ldots q_T$.

3. Given a number of observations, learn the optimal values of the parameters of $\lambda = (A,B,C,\pi)$ to maximize $P(O|\lambda)$ for all the observations.

For a detailed tutorial on how HMMs solve these problems the reader is referred to Rabiner [10]. Here, the model parameters are as follows: A is a 2D matrix holding the elements a_{ij}= Probability of going from state $q_t = S_i$ to $q_{t+1} = S_j$ for all $1 \leq i,j \leq N$, N is the total number of states; B is a 2D matrix holding the elements b_{jk} = probability of observation $O_t = V_k$ given you are in state j and $0 \leq k \leq M$ (total number of possible observations is M); C is a 3D matrix holding $c_i(h,k)$ = probability of observing $O_{t+1} = V_k$ given we are in state $q_t = S_i$, observing $O_t = V_h$; π is a vector of π_i = initial probability of being in state $q_1 = S_i$.

To solve the first problem, we develop a forward algorithm procedure to compute $\alpha_i(t) = p(O_1,O_2,\ldots O_t,q_t = i|\lambda)$. When t = T, $P(O|\lambda)$ is found by summing all the α_i's at time T. The forward algorithm procedure is:

1. $\alpha_i(1) = \pi_i b_i(O_1)$ for all i, $0 \leq i, t \leq T$, and $b_i(O_1) = b_{ih}$ for some h which $O_1 = V_h$.

2. $\alpha_j(t+1) = \left[\sum_{i=1}^{n} \alpha_i(t)\alpha_{ij}c_i(O_t, O_{t+1})\right]b_j(O_{t+1})$, where $c_i(O_t,O_{t+1})$ is $c_i(h,k)$ for $O_t = V_h$ and $O_{t+1} = V_k$ and n is the total number of hidden states.

3. $p(O|\lambda) = \sum_{i=1}^{n} \alpha_i(T)$.

The final probability $p(O|\lambda)$ is the probability we are looking for.

A backwards algorithm procedure can also be developed to find $P(O|\lambda)$. The variable β_i must be created such that $\beta_i(t) = p(O_{t+1},O_{t+2},\ldots,O_T \mid q_t = i, \lambda)$.

1. $\beta_i(T) = 1$

2. $\beta_i(t) = \left[\sum_{j=1}^{n} a_{ij}b_j(O_{t+1})\,\beta_j(t+1)\right]c_i(O_t, O_{t+1})$

$p(O|\lambda) = \sum_{i=1}^{n} \beta_i(1)\,\pi_i b_i(O_1)$.

It should be noted that the probability of the observations given the model using both the forward and backwards algorithms are used later to help find answers to the remaining two Evidence Feed-Forward HMM problems.

To solve the second problem, computing the optimal path of hidden states from the observations, given the model, one must make use of both the backwards and forwards algorithm. Optimal path is assumed that we are looking for the path that gives the maximum probability of the state sequence given the observations and the model. We are maximizing $P(Q|O,\lambda)$. To do this we create two new variables, δ and Path.

1. $\delta_1(i) = \pi_i b_i(O_1)$. Path = [].

2. $\delta_t(j) = \max_{1 \leq i \leq n}\left[(\delta_{t-1}a_{ij})b_j(O_t)c_i(O_{t-1}, O_t)\right]$. Path is state which this is maximized. Add the state to the Path.

3. Final step is finding the state which maximizes $\delta_T(i)$ for $1 \leq i \leq n$.

To solve the final problem, we use the Baum-Welch algorithm to optimize the parameters. First the equation is separated into four parts and a constraint is applied.

$$\sum_{i=1}^{N} \pi_i = 1,$$

$$\sum_{j=1}^{N} a_{ij} = 1,$$

$$\sum_{k=1}^{M} b_{jk} = 1 \text{ for all j,}$$

$$\sum_{k=1}^{M} C_i(h,k) = 1,$$

Next, create the variables $\gamma_i(t) = p(q_t=i|O,\lambda)$, the probability of being in state i at time t for sequence O and model λ, and the variable $\xi_{ij}(t)=P(q_t=1, q_{t+1} = j)|O,\lambda)$, the probability of being in state i at time t and state j at time t+1 given the observations and the model.

Using LaGrange and the constraints above, we end up with the re-estimated parameters as:

$\overline{\pi_i}$ = expected number of times in state i at time t=1,
$\overline{\pi_i} = \gamma_i(1)$,

$\overline{a_{ij}}$ = (expected number of transitions from state i to state j) / (expected number of transitions from state i)

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \xi_{ij}(y)}{\sum_{t=1}^{T-1} \gamma_i(t)},$$

$\overline{b_{ik}}$ = (Expected number of times in state j and observing O_k) / (expected number of times in state j)

$$\overline{b_{ik}} = \frac{\sum_{t=1}^{T} \gamma_j(t), where\ O_t = O_k}{\sum_{t=1}^{T} \gamma_j(t)},$$

$\overline{C_i(h,k)}$ = (Expected number of times in state i observing O_h and transitioning observing O_k in state j for all j) / expected number of times in state i observing O_h).

$$\overline{c_i(h,k)} = \frac{\sum_{t=1}^{T-1} \gamma_i, where\ O_t = O_h\ and\ O_{t+1} = O_k}{\sum_{t=1}^{T-1} \gamma_i, where\ O_t = O_h}.$$

5. RESULTS

Data pertaining to some common activities (jump, jog, dribble a basketball, kick a soccer ball) were simulated. The data was over-processed to reduce most all common features that would normally be used to detect the activity. The reason for using the over-processed data is to show that, 1.) the Evidence Feed-Forward HMM can still detect patterns from messy, sparse data; and 2.) The Evidence Feed-Forward HMM would obviously show a better detection rate when compared with standard HMMs.

First, the simulated images were extracted from the video, ranging from an activity of 20 frames to an activity of over 100 frames. Next, each frame was processed to detect the hands, feet, and head. A single point represented each hand, each foot and the head. Often times these points were mis-represented. A bounding box was put around the five points and the height of the bounding box was divided by the width. This number was put into 1 of 11 bins equally divided between the highest and lowest value. A graph of the bin values for each frame can be found in figure 3.

Finally the bin values for each frame was compared to its next bin value and a symbol associated with increase, decrease, or stay the same was used. This symbol was the input parameter for both the Evidence Feed-Forward HMM and the standard HMM.

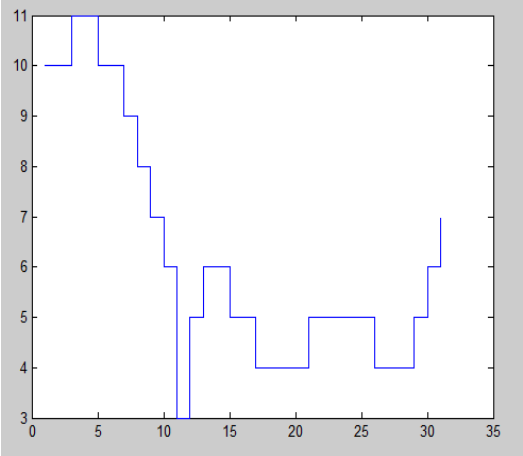

Fig. 3. Processed data for a sequence representing SOCCER KICK. The x values represent the frame in the sequence. The y values represent the bin the height/width ratio belongs to.

Both HMMs were trained with only four activity sequences for the JUMP, JOG, and DRIBBLE activities. The SOCCER KICK activity had only three sequences trained. Of these activity's a testing set was used that did not include the training set. For the results for the JUMP activity, the Evidence Feed-Forward HMM correctly classified 78% where the standard HMM classified 67% correctly. For the SOCCER KICK activity, the Evidence Feed-Forward HMM classified 100% correctly, where the standard HMM classified 50% correct. The DRIBBLE activity saw results of 12.5% for the HMM and 50% for the Evidence Feed-Forward HMM. Finally, the JOG activity did not fair so well for any of the classifiers. The HMM classified JOG correctly 21% of the time. For the Evidence Feed-Forward HMM, 46% of the sequences were correctly classified.

6. CONCLUSION

This paper has shown the idea and theory behind the Evidence Feed Forward Hidden Markov Model.

The idea behind building this HMM is based around the assumption of the observations being affected by the previous observation. This is not the case in standard HMMs. Adding this probability to the classifier improves the classification rate greatly on messy, sparse data sets, as shown in the results section. To tackle the complex problems associated with Visual Understanding, a more complex technique needs to be developed. It is the hope of this paper to convince the reader that the Evidence Feed-Forward HMM is one such technique. Further studies on sparse, messy data is to be investigated in the near future.

7. REFERENCES

[1] Cristani, M., Bicego, M., Murino, V., "Audio-Visual Event Recognition in Surveillance Video Sequences," *IEEE Transactions on Multimedia*, Vol. 9, No. 2, Feb. 2007, pp. 257-267.

[2] Dimitrijevic, M., Lepetit, V., Fua, P., "Human Body Pose Detection Using Bayesian Spatio-Temporal Templates," *Computer Vision and Image Understanding*, Vol. 104, No. 2, 2006, pp.127-139.

[3] Stern, H., Kartoun, U., Shmilovici, A., "A Prototype Fuzzy System for Surveillance Picture Understanding," *Proceedings from Visual Imaging and Image Processing Conference*, Sept 2001, pp. 624-629.

[4] Ogale, A., Karapurkar, A., Aloimonos, Y., "View-Invariant Modeling and Recognition of Human Actions Using Grammars," *Lecture Notes in Computer Science*, Vol. 4358, Springer, Berlin, 2007, pp.115-126, doi:10.1007/978-3-540-70932-9_9

[5] Yamato, J., Ohya, J., Ishii, K., "Recognizing Human Action in Time Sequential Images Using Hidden Markov Models," *Proceedings from IEEE Computer Vision and Pattern Recognition (CVPR)*, 1992, pp. 379-385.

[6] Wilson, A., Bobick, A., "Parametric Hidden Markov Models for Gesture Recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 21, 1999, pp. 884-899.

[7] Oliver, N. M., Rosario, B., Pentland, A. P., "A Bayesian Computer Vision System for Modeling Human Interaction," *IEEE Transactions in Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, Aug. 2000, pp. 831-843.

[8] Xiang, T., Gong, S., "Activity Based Video Content Trajectory Representation and Segmentations," *Proceedings from British Machine Vision Conference*, 2004, pp. 177-186.

[9] Xiang, T., Gong, S., "Incremental Visual Behaviour Modelling," *Proceedings from European Conference on Computer Vision*, 2006, pp. 65-72.

[10] Rabiner, L., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.", Proceedings of IEEE, Vol. 77(2), pp.257-286.